

Probability density of quantum expectation values

L Campos Venuti and P Zanardi

Department of Physics and Astronomy and Center for Quantum Information Science & Technology, University of Southern California, Los Angeles, California 90089-0484, USA

E-mail: camposv@usc.edu

Abstract. We consider the quantum expectation value $\mathcal{A} = \langle \psi | A | \psi \rangle$ of an observable A over the state $|\psi\rangle$. We derive the exact probability distribution of \mathcal{A} seen as a random variable when $|\psi\rangle$ varies over the set of all pure states equipped with the Haar-induced measure. The probability density is obtained with elementary means by computing its characteristic function, both for non-degenerate and degenerate observables. To illustrate our results we compare the exact predictions for few concrete examples with the concentration bounds obtained using Levy's lemma. Finally we comment on the relevance of the central limit theorem.

1. Introduction

The role of probability distributions in quantum theory cannot be overestimated. Arguably the most important of those distributions is the one describing the statistics of possible outcomes of the measurement of the observable associated with the self-adjoint operator A while the system is in the pure state $|\psi\rangle$: $P_A(a) := \langle \psi | \delta(A - a) | \psi \rangle$. This function is supported on the numerical range of A and, for bounded A can be equivalently characterized by the set of its moments (see e.g. [1]): $m_k := \langle \psi | A^k | \psi \rangle$ ($k \in \mathbb{N}$), i.e., the expectation values of the family of observables $\{A^k\}_{k \in \mathbb{N}}$ in the state $|\psi\rangle$. This latter, quite often, can be itself regarded as a random variable distributed according to some prior density that depends on the problem under consideration. For example, in the context of equilibration dynamics of closed quantum systems [2, 3] one is interested in the quantity $a(t) := \langle \psi(t) | A | \psi(t) \rangle$, where $|\psi(t)\rangle := e^{-itH}|\psi\rangle$ and H is the Hamiltonian operator of the system. If one monitors A by sampling time instants uniformly over the interval $[0, T]$ the underlying probability space for the $|\psi(t)\rangle$'s is the segment $[0, T]$ equipped with the uniform measure dt/T . In this case averaging over the quantum states amounts to perform the time average $1/T \int_0^T a(t) dt$.

Another possibility that recently gained relevance in the foundation of statistical mechanics [4] is to consider $\langle \psi | A | \psi \rangle$ and let $|\psi\rangle$ vary over the full unit sphere of the, say d -dimensional, Hilbert space space. This manifold is transitively acted upon by the group of all $d \times d$ unitary matrices $\mathbb{U}(d)$ and therefore inherits a natural invariant measure from the unique group-theoretic invariant measure over $\mathbb{U}(d)$ i.e., the Haar measure.

In this paper we will address precisely this latter setting and compute the probability distribution for the quantum expectations $\langle \psi | A^k | \psi \rangle$ seen as random variables over the unit sphere of the Hilbert space equipped with the measure induced by the Haar measure. In this way the function P_A itself becomes a sort probability-density valued random variable that can be partially characterized by the probability densities of its moments m_k over the unit sphere. We will show

that these probability densities can be determined with elementary tools and explicit analytical expressions for their characteristic functions can be obtained.

After the completion of this work we became aware that similar results are contained in the recent works [5, 6, 23]. See also [7] for an entry into the mathematical literature.

2. Preliminaries

Our key object is $\mathcal{A}(\psi) = \langle \psi | A | \psi \rangle$. Throughout this paper we write $\overline{f(\psi)} = \int D\psi f(\psi)$ for Haar-induced averages over pure states. Computing the first few moments of $\mathcal{A}(\psi)$ is a relatively easy task. The first moment reads

$$m_1 = \overline{\mathcal{A}} = \int D\psi \operatorname{tr}(A|\psi\rangle\langle\psi|) = \frac{\operatorname{tr}(A)}{d}. \quad (1)$$

a result which follows from $\overline{|\psi\rangle\langle\psi|} = \mathbb{I}/d$ [8]. A closed formula for the general moment can be obtained by noting that

$$\overline{|\psi\rangle\langle\psi|^{\otimes n}} = \frac{1}{\binom{d+n-1}{n}} \frac{1}{n!} \sum_{\pi \in S_n} P_\pi. \quad (2)$$

Here P_π is the operator that enacts the permutation π in $\mathcal{H}^{\otimes n}$ and S_n is the symmetric group of n elements. For a proof of (2) see e.g. [9]. The proportionality constant is obtained noting that $\binom{d+n-1}{n}$ is the dimension of the totally symmetric space, and the remaining operator is an orthogonal projector. Using eq. (2) one obtains the following closed expression for the n -th moment

$$\begin{aligned} m_n &= \frac{(d-1)!}{(d+n-1)!} \sum_{\pi \in S_n} \operatorname{tr}(P_\pi A^{\otimes n}) \\ &= \frac{(d-1)!}{(d+n-1)!} \sum_{\pi \in S_n} \sum_{\sigma_1=1}^d \cdots \sum_{\sigma_n=1}^d A_{\sigma_1, \sigma_{\pi(1)}} \cdots A_{\sigma_n, \sigma_{\pi(n)}}. \end{aligned} \quad (3)$$

This expression is a sum of contractions and represents a sum of products of traces of A . For instance for $n = 2, 3$ one has

$$m_2 = \frac{\operatorname{tr}(A)^2 + \operatorname{tr}(A^2)}{d(d+1)}, \quad m_3 = \frac{\operatorname{tr}(A)^3 + 3 \operatorname{tr}(A^2) \operatorname{tr}(A) + 2 \operatorname{tr}(A^3)}{d^3 + 3d^2 + 2d}. \quad (4)$$

In any case it seems difficult to obtain the probability density with this approach. Instead our procedure will be that of computing directly the characteristic function $\chi(\lambda) := \overline{e^{i\lambda\mathcal{A}(\psi)}}$ and obtain the probability density by Fourier transforming.

Choosing a basis $|j\rangle$, ($j = 1, \dots, d$) and calling $z_j = \langle j | \psi \rangle$ we observe that we can write the average over $|\psi\rangle$ as

$$\overline{f(\psi)} = C \int \delta\left(\sum_{j=1}^d |z_j|^2 - 1\right) f(\psi) d^2 \mathbf{z} \quad (5)$$

where we defined

$$\int d^2 \mathbf{z} = \prod_{i=1}^d \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{dx_i dy_i}{\pi}.$$

The normalization constant C can be computed with the same technique that we are going to show and it turns out to be equal to $(d-1)!$. Using the Fourier representation for the delta function in eq. (5) we obtain a Gaussian integral that can be computed. For simplicity we use the basis that diagonalizes A : $A = \sum_j a_j |j\rangle\langle j|$. Calling $D_A = \text{diag}\{a_1, a_2, \dots, a_d\}$ we can write the characteristic function as

$$\chi(\lambda) = (d-1)! \int d^2 \mathbf{z} \int_{-\infty}^{\infty} \frac{dr}{2\pi} e^{ir(z^\dagger z - 1)} e^{i\lambda z^\dagger D_A z - \epsilon z^\dagger z}.$$

As customary we introduced a small positive ϵ in order to make the Gaussian integral absolutely convergent. The Gaussian integration gives

$$\chi(\lambda) = \frac{(d-1)!}{(-i)^d} \int_{-\infty}^{\infty} \frac{dr}{2\pi} \frac{e^{-ir}}{\prod_j (r - r_j)}, \quad r_j = -\lambda a_j - i\epsilon \quad (6)$$

At this point we make the important assumption that all eigenvalues of A are non-degenerate. We will treat the general case in section 3. Under these conditions the integrand in eq. (6) has only simple poles and the integral is easily evaluated with residues closing the circle in the lower half-plane. The result is, after sending $\epsilon \rightarrow 0$,

$$\chi(\lambda) = \frac{(d-1)!}{(i\lambda)^{d-1}} \sum_{k=1}^d \frac{e^{i\lambda a_k}}{\prod_{j \neq k} (a_k - a_j)}. \quad (7)$$

Although it might not be readily apparent from eq. (7), $\chi(\lambda)$ is actually regular in $\lambda = 0$. This fact follows from a set of identities proven in Appendix A stating that

$$\sum_{k=1}^d \frac{(a_k)^n}{\prod_{j \neq k} (a_k - a_j)} = \begin{cases} 0 & 0 \leq n \leq d-2 \\ 1 & n = d-1 \end{cases}. \quad (8)$$

Applying eq. (8) to eq. (7) we thus see that $\chi(\lambda)$ is regular at $\lambda = 0$ and being a linear combination of analytic functions it is in fact analytic in the whole complex plane (entire). A simple way to remember that $\chi(\lambda)$ must be regular at $\lambda = 0$ is to note that $\chi(\lambda) = 1 + m_1(i\lambda) + O(\lambda^2)$. In fact this very same approach can be used to prove eq. (8). Using eq. (8) and eq. (7) we readily obtain the Taylor series of $\chi(\lambda)$

$$\chi(\lambda) = \sum_{n=0}^{\infty} m_n \frac{(i\lambda)^n}{n!} \quad (9)$$

$$m_n = \frac{1}{\binom{n+d-1}{n}} \sum_{k=1}^d \frac{(a_k)^{n+d-1}}{\prod_{j \neq k} (a_k - a_j)}. \quad (10)$$

Equation (10) is a quite compact expression in place of the complicated eq. (3). Equating eq. (10) with eq. (3) we obtain the following non-trivial set of matrix identities valid when the spectrum of A is non-degenerate

$$\frac{1}{n!} \sum_{\pi \in S_n} \text{tr}(P_\pi A^{\otimes n}) = \sum_{k=1}^d \frac{(a_k)^{n+d-1}}{\prod_{j \neq k} (a_k - a_j)}.$$

Since $\chi(\lambda)$ is entire, analytic continuation is trivial and the moment generating function $M(y) := \overline{e^{y\bar{A}}}$ can be simply obtained by setting $i\lambda = y$ in eq. (7).

To obtain the probability density we must compute the Fourier transform of χ

$$P(x) = \int_{-\infty}^{\infty} \frac{d\lambda}{2\pi} e^{-ix\lambda} \chi(\lambda). \quad (11)$$

We make use of the following trick. Since $\chi(\lambda)$ is well behaved in zero and decays at infinity, it belongs to the space of tempered distribution $\mathcal{S}'(\mathbb{R})$. The Fourier transform is well defined for such functions and the result is again a tempered distribution. Actually for $d \geq 3$, χ is also summable and so its Fourier transform can be safely defined by the absolutely converging integral eq. (11). We proceed then forgetting about the behavior in $\lambda = 0$ and compute the Fourier transform as a linear combination of Fourier transform of the functions $e^{i\lambda a_k} / (i\lambda)^{d-1}$. Such functions however are not well behaved in zero (they are not distributions) and must be regularized in order to compute the Fourier transform. Since the total Fourier transform is well defined, the result cannot depend on the regularization. A convenient regularization of $1/\lambda^n$ for n integer, is to be proportional to the n -th derivative of $\ln|\lambda|$ (" λ^{-n} " := $(-1)^{n-1} \partial_\lambda \ln|\lambda| / (n-1)!$). With this definition the Fourier transform, in the sense of distribution, is

$$\int_{-\infty}^{\infty} \frac{d\lambda}{2\pi} \frac{e^{i\lambda y}}{(i\lambda)^n} = \frac{y^{n-1} \text{sign}(y)}{2(n-1)!}.$$

Using this equation and eq. (7) we obtain

$$P(x) = \frac{(d-1)}{2} \sum_k \frac{(a_k - x)^{d-2} \text{sign}(a_k - x)}{\prod_{j \neq k} (a_k - a_j)}. \quad (12)$$

By construction, $P(x)$ has all the properties of a probability density, moreover, since the probability of obtaining a value outside the numerical range of A must be zero, $P(x)$ must be supported in $[\min\{a_k\}, \max\{a_k\}]$ which, once again, is not readily apparent from eq. (12). That $P(x)$ is compactly supported follows from Paley-Wiener theorem since it is the Fourier transform of an $L^2(\mathbb{R})$ analytic function. In any case a direct proof that $P(x) = 0$ for x outside $[\min\{a_k\}, \max\{a_k\}]$ is given in Appendix A and follows directly from eq. (A.3).

Generically $P(x)$ is a bell-shaped curve supported in $[\min\{a_k\}, \max\{a_k\}]$ as can be see from fig. 1. The smoothness properties of $P(x)$ can be understood noting that $\partial_x^{(d-2)} P(x)$ is a piecewise constant function (a part from a bunch of deltas).

2.1. Cumulative distribution function

Having computed the probability density, the characteristic, and the moment generating function, we would like now to compute the cumulative distribution function (CDF) to complete the picture. The CDF can be obtained via the following integral

$$\text{cdf}(x) = \text{Prob}(\mathcal{A} \leq x) = \int_{-\Omega}^x P(t) dt,$$

as long as $-\Omega < \min\{a_k\}$. The integration of each term in eq. (12) gives

$$\int_{-\Omega}^x (t - a_k)^{d-2} \text{sign}(t - a_k) dt = \text{sign}(x - a_k) \frac{(x - a_k)^{d-1}}{(d-1)} + \frac{(-\Omega - a_k)^{d-1}}{(d-1)}.$$

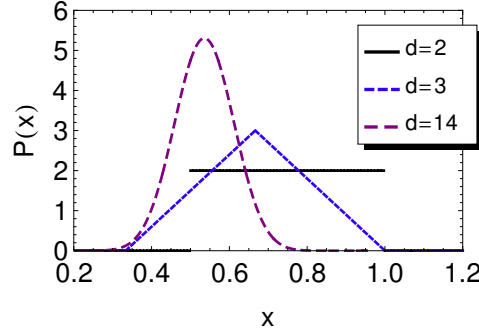


Figure 1. Plot of the probability density eq. (12) for various d . In this particular case we set $a_k = k/d$.

Clearly the result cannot depend on Ω as long as $-\Omega < \min\{a_k\}$. In fact, using again formula (A.3) of Appendix A, the sum of all the terms containing Ω gives

$$\frac{1}{2} \sum_{k=1}^d \frac{(-\Omega - a_k)^{d-1}}{\prod_{j \neq k} (a_j - a_k)} = \frac{1}{2}. \quad (13)$$

All in all the final expression for the cumulative is

$$\text{cdf}(x) = \frac{1}{2} + \frac{1}{2} \sum_k \text{sign}(x - a_k) \frac{(x - a_k)^{d-1}}{\prod_{j \neq k} (a_j - a_k)} \quad (14)$$

3. General case

So far we assumed that the spectrum of A was non-degenerate. Although this is the most common situation, one may want to consider the case where A is a projector. We then turn to the fully general case where the spectrum of A consists of ℓ distinct eigenvalues $\{a_j\}_{j=1}^\ell$ with degeneracy n_j satisfying $\sum_{j=1}^\ell n_j = d$. For a change, let us compute the moment generating function $M(y, \Omega) := \overline{e^{y(A-\Omega)}}$ instead of the characteristic function. The shift constant is chosen to satisfy $\Omega > \max\{a_j\}$ so that for $y > 0$ the Gaussian integral is well defined. Gaussian integration gives the same result as for the non-degenerate case, which we now re-write as

$$M(y, \Omega) = \frac{(d-1)!}{(-i)^d} \int_{-\infty}^{\infty} \frac{dr}{2\pi} \frac{e^{-ir}}{\prod_{j=1}^\ell (r - r_j)^{n_j}}. \quad (15)$$

To compute the above integral with the residues we must compute the following derivatives

$$\text{Res} \left[\frac{e^{-ir}}{\prod_{j=1}^\ell (r - r_j)^{n_j}}, r = r_k \right] = \frac{1}{(n_k - 1)!} \partial_r^{(n_k-1)} \left[\frac{e^{-ir}}{\prod_{j \neq k} (r - r_j)^{n_j}} \right] \Big|_{r=r_k} \quad (16)$$

Alternatively, another possibility is to use the Chinese Remainder Theorem (see e.g. [10])[‡] to get the partial fraction decomposition of the function in eq. (15). This way one expresses that function

[‡] LCV wishes to thank Pierre-Yves Gaillard for pointing out this connection.

as a sum of functions with simple poles so that the residues can be evaluated straightforwardly. The results coincide and in fact eq. (16) provides a way to obtain the partial fraction decomposition of the function in (15) by differentiation. The procedure is outlined in A.1, and the final result for the moment generating function is

$$M(y, \Omega) = (d-1)! \sum_{k=1}^{\ell} \sum_{M_k=0}^{n_k-1} \frac{(-1)^{M_k} e^{y(a_k-\Omega)}}{y^{d+M_k-n_k} (n_k-1-M_k)!} \times \\ \times \sum_{\sum_{j \neq k}^{\ell} m_j = M_k} \prod_{\substack{j=1 \\ j \neq k}}^{\ell} \frac{(n_j+m_j-1)!}{m_j! (n_j-1)!} \frac{1}{(a_k-a_j)^{n_j+m_j}} \quad (17)$$

which correctly reduces to eq. (A.1) when all $n_k = 1$. As for the non-degenerate case, this function is regular at $y = 0$, and it is fact an entire function. The probability density can be obtained with the same arguments used for the non-degenerate case and the result is

$$P(x) = \sum_{k=1}^{\ell} \sum_{M_k=0}^{n_k-1} \frac{(a_k-x)^{d+M_k-n_k-1} \text{sign}(a_k-x) (-1)^{M_k} (d-1)!}{2(d+M_k-n_k-1)! (n_k-1-M_k)!} \times \\ \sum_{\sum_{j \neq k}^{\ell} m_j = M_k} \prod_{\substack{j=1 \\ j \neq k}}^{\ell} \binom{n_j+m_j-1}{m_j} \frac{1}{(a_k-a_j)^{n_j+m_j}}. \quad (18)$$

As a concrete example we will consider in some detail the case where A is a one-dimensional projector.

3.1. Probability density for the random guess

The case where A is a one-dimensional projector is of some relevance and we discuss it with some detail. Because of unitary invariance of the measure we can fix the reference state to be $|1\rangle$ in some basis, i.e. $A = |1\rangle\langle 1|$. In this case the quantity \mathcal{A} that we are considering is the fidelity between a reference state $|1\rangle$ and a random vector $|\psi\rangle$, i.e. $\mathcal{A} = |\langle 1|\psi\rangle|^2$ (normally written as \mathcal{F}). It is a standard textbook exercise (see e.g. [11] chapter two) to compute the average random guess. Using eq. (1) one gets $\overline{\mathcal{F}} = 1/d$, but what is the form of the probability distribution? Since A is a one-dimensional projector it has two eigenvalues $a_1 = 1$ and $a_2 = 0$ with multiplicity $n_1 = 1$ and $n_2 = d-1$ respectively. Inserting these values in eq. (17) we obtain for the moment generating function

$$M_{\mathcal{F}}(y, \Omega) = (d-1)! \left\{ \frac{e^{y(1-\Omega)}}{y^{d-1}} - \sum_{m=0}^{d-2} \frac{e^{-y\Omega}}{y^{m+1} (d-2-m)!} \right\}. \quad (19)$$

Once again, despite its appearance, this function is regular at $y = 0$ and in fact entire. Rearranging the terms in the sum we can write it as

$$M_{\mathcal{F}}(y, \Omega) = e^{-\Omega y} \sum_{n=0}^{\infty} y^n \frac{(d-1)!}{(d+n-1)!}.$$

Setting $\Omega = 0$ we readily obtain the moments as

$$\overline{\mathcal{F}^n} = \frac{n! (d-1)!}{(d+n-1)!}, \quad (20)$$

which agrees with a result of Von Neumann [12, 13]. The probability density can be obtained either by Fourier transforming eq. (19) or by setting $a_1 = 1$, $n_1 = 1$ and $a_2 = 0$, $n_2 = d - 1$ in eq. (18). The sum over M_2 becomes a binomial and the result is surprisingly simple

$$P_{\mathcal{F}}(x) = (d-1)(1-x)^{d-2} \mathbb{I}_{[0,1]}(x). \quad (21)$$

Here we denoted by $\mathbb{I}_{[0,1]}(x)$ the indicator function of the set $[0,1]$, i.e. $\mathbb{I}_{[0,1]}(x) = [\text{sign}(x) + \text{sign}(1-x)]/2$. Eq. (21) is the so-called beta distribution with parameters $\alpha = 1$ and $\beta = d - 1$. In Bayesian statistics, the beta distribution can be seen as the posterior probability of the parameter p of a binomial distribution after observing $\alpha - 1$ successes (with probability of success given by p) and $\beta - 1$ failures (with probability of failure $1 - p$). The CDF is given by a simple integration and reads

$$\text{cdf}_{\mathcal{F}}(x) = 1 - (1-x)^{d-1}, \quad \text{for } 0 \leq x \leq 1, \quad (22)$$

whereas $\text{cdf}_{\mathcal{F}}(x) = 0$, (1) for $x < 0$ ($x > 1$) respectively.

4. Comparison with Levy's lemma

A typical approach to gain information on the concentration properties of a random variable X , is to compute the first few moments of the variable and then use variations of Markov's or Chebyshev's inequalities to obtain a bound on $\text{Prob}(X - \bar{X} > \epsilon)$. Alternatively in some cases one can use the Levy's lemma which typically provides tighter bounds. Since we obtained the probability density exactly we would like to compare the exact concentration prediction with that obtained by Levy's bound.

Roughly speaking Levy's lemma states that, for a vector in a large-dimensional hypersphere the probability that a (sufficiently smooth) function is far from its average is exponentially small in the dimension of the space. There are many versions of Levy's lemma involving either the mean or the median or differing for the definition of being "far from". To be specific we use the following version [14]: for $x \in \mathbb{S}^k = \{y \in \mathbb{R}^{k+1}, \|y\| = 1\}$, and for a function f with Lipschitz constant η

$$\text{Prob}\{f(x) - \langle f \rangle \geq \epsilon\} \leq 2 \exp(-C_1(k+1)\epsilon^2/\eta^2) \quad (23)$$

with $C_1 = (9\pi^3 \ln 2)^{-1}$. In our setting $|\psi\rangle$ lives in a d dimensional complex space so $k = 2d - 1$.

We will compare the prediction of the Levy's lemma to our exact result for two concrete examples of operator A .

4.1. Random guess

First we consider again the fidelity of the random guess, in which case A is a one-dimensional projector. As we mentioned previously, the average fidelity is $\bar{\mathcal{F}} = 1/d$, so to compute the LHS of eq. (23) it suffices to notice that $\text{Prob}(\mathcal{F} > x) = 1 - \text{cdf}_{\mathcal{F}}(x)$ and set $x = \bar{\mathcal{F}} + \epsilon = 1/d + \epsilon$ in eq. (22). Then

$$\begin{aligned} \text{Prob}(\mathcal{F} - \bar{\mathcal{F}} \geq \epsilon) &= \left(1 - \frac{1 + \epsilon d}{d}\right)^{d-1} \\ &= \frac{e^{-(d-1)/(d(1-\epsilon))}}{1 - \epsilon} e^{\ln(1-\epsilon)d} \\ &\simeq e^{-1} e^{-\epsilon d}, \end{aligned} \quad (24)$$

where the last equation has been obtained in the limit of ϵ small and d large. The Lipschitz constant of the function $\mathcal{A}(\phi) = \langle \phi | A | \phi \rangle$ has been calculated in [15] Appendix A (see also [4]) where it has been shown that $|\mathcal{A}(\phi_1) - \mathcal{A}(\phi_2)| \leq 2 \|A\|_{op} \|\phi_1 - \phi_2\|$ where $\|A\|_{op}$ is the operator norm of A (maximum singular value). In our case $A = |1\rangle\langle 1|$ so $\|A\|_{op} = 1$ and we may take $\eta = 2$ in eq. (23) above. So the Levy's lemma predicts

$$\text{Prob}(\mathcal{F} - \overline{\mathcal{F}} \geq \epsilon) \leq e^{-C\epsilon^2 d}$$

$$C = \frac{1}{18\pi^3 \ln 2} \simeq \frac{1}{387}$$

As we see comparing with eq. (24) Levy's bound pays a very small pre-factor in the exponential compared to the exact behavior and a quadratic rather than a linear dependence on the error ϵ .

4.2. Number operator

Let us now take an example from the non-degenerate case and take A to be the number operator \hat{N} , i.e. the operator with $a_k = k$ for $k = 1, 2, \dots, d$. For this operator we have $\overline{\mathcal{A}} = \text{tr } A/d = (d+1)/2$. Since the norm of \hat{N} grows linearly with d we do not expect concentration to take place in this case. Now

$$\text{Prob}\{\mathcal{A} - \overline{\mathcal{A}} \geq \epsilon\} = 1 - \text{cdf}\left(\frac{d+1}{2} + \epsilon\right) =: B(d, \epsilon)$$

We investigated numerically the function $B(d, \epsilon)$ using eq. (14). Our results indicate that for large d and small ϵ , $B(d, \epsilon) \simeq \alpha e^{-C\epsilon}$ where the constant $C \approx 0.25$. The operator norm is $\|\hat{N}\|_{op} = d$ so Levy's lemma in this case predicts a bound independent of d :

$$B(d, \epsilon) \leq 2e^{-C'\epsilon^2}$$

$$C' = \frac{2}{9\pi^3 \ln 2} \simeq 0.0103 \dots \simeq \frac{1}{97}.$$

From this comparison we can draw similar conclusions as for the case of the random guess. Namely Levy's lemma predicts a qualitatively correct behavior in d , but with a much larger constant: $C \approx 24C'$. The quadratic dependence on the error is typically smaller and in both cases seen here, it was seen in fact to be linear for small enough ϵ .

5. Central limit theorem?

If the (non-degenerate) self-adjoint operator A has the following spectral resolution $A = \sum_{k=1}^d a_k |k\rangle\langle k|$, our random variable $\mathcal{A}(\psi)$ can be written as a weighted sum of the random variables $X_k(\psi) := |\langle \psi | k \rangle|^2$. In the limit of large Hilbert space dimension d it is easy to see that the X_k 's decouple. Indeed if $h \neq k$ one has

$$\begin{aligned} \overline{X_k X_h} &= \overline{\langle \psi | k \rangle \langle k | \psi \rangle \langle \psi | h \rangle \langle h | \psi \rangle} = \overline{\text{tr} [|\psi\rangle\langle\psi|^{\otimes 2} |k\rangle\langle k| \otimes |h\rangle\langle h|]} \\ &= \text{tr} \left[\frac{1+P}{d(d+1)} |k\rangle\langle k| \otimes |h\rangle\langle h| \right] = \frac{1}{d(d+1)}, \end{aligned} \quad (25)$$

while for $h = k$ $\overline{X_k^2} = 2/d(d+1)$ from eq. (20) with $n = 2$. Whence one obtains for the covariance:

$$\text{cov}(X_h, X_k) := \overline{X_k X_h} - \overline{X_k} \overline{X_h} = \frac{1}{d(d+1)} - \frac{1}{d^2} = O(1/d^3),$$

for $h \neq k$ and $\text{cov}(X_k, X_k) = O(2/d^2)$, i.e. the off-diagonal terms decay faster than the diagonal ones. This fact is a consequence of a well known, general result. In physics a manifestation of this phenomenon is the fact that the large- N limit of $O(N)$ invariant field theories (such as the non-linear sigma model) is a free, Gaussian, theory (see e.g. [16]). Probabilists generally tend to see this result in the opposite direction, namely: in the large N limit, the Gaussian measure becomes highly concentrated on a hyper-sphere of given radius. Given these considerations it is natural to expect that the variable \mathcal{A} satisfies a central limit theorem (CLT), at least for a certain class of operators A . By CLT we mean here that the rescaled random variable

$$\mathcal{Z} := (\mathcal{A} - \overline{\mathcal{A}}) / \sqrt{\kappa_2(\mathcal{A})},$$

($\kappa_n(\mathcal{A})$ n -th cumulant) tends in distribution to a Gaussian with zero mean and unit variance as $d \rightarrow \infty$.

Instead of trying to give a complete characterization of the precise conditions on A for the CLT to apply we will content here with a few examples (see however [7]). Consider first the class of A with $a_k = k^\alpha$ (with say $\alpha > 0$). Using formulae (1) and (4) we obtain, at leading order $\kappa_1(\mathcal{A}) \propto d^\alpha$, $\kappa_2(\mathcal{A}) \propto d^{2\alpha-1}$, and $\kappa_3(\mathcal{A}) \propto d^{3\alpha-2}$ from which we get $\kappa_3(\mathcal{Z}) \propto 1/\sqrt{d}$, i.e. the third cumulant of \mathcal{Z} vanishes for large d . The approach to Gaussian for this case can be read off from figure 2. A potentially interesting class of operators consists of those A for which the average scales as the volume of the space i.e. the log of the Hilbert's space dimension. As a last example we consider then $a_k = \log k$. Correctly in this case one has $\overline{\mathcal{A}} \propto \log d$. Proving that the third cumulant of \mathcal{Z} goes to zero is now more complicated. However Euler-Maclaurin formula is sufficient to prove the following result valid at leading order

$$\begin{aligned} \kappa_2(\mathcal{A}) &= \frac{\sum_{k=1}^d (\ln(k) - \frac{\overline{x}}{d})^2}{d(d+1)} = \frac{1}{d} + \text{smaller terms} \\ \kappa_3(\mathcal{A}) &= \frac{2 \sum_{k=1}^d (\ln(k) - \frac{\overline{x}}{d})^3}{d(d+1)(d+2)} = \frac{-4}{d^2} + \text{smaller terms} \end{aligned}$$

where $\overline{x} = \sum_{k=1}^d \ln k = \ln(d!)$. Hence we see that also in this case $\kappa_3(\mathcal{Z}) \propto 1/\sqrt{d}$. The approach to Gaussian is depicted in figure 2.

6. Conclusions

In this paper we considered the quantum expectation value of an operator A with respect to a pure state $|\psi\rangle$. We computed exactly the probability density of the expectation value when $|\psi\rangle$ is drawn from the space of pure states according to the unique (Haar-induced) unitarily invariant measure. Generically, the resulting probability distribution is a compactly supported, piecewise polynomial function. We used the exact result to test the tightness of the concentration bounds obtained by Levy's lemma for a couple of particular cases, namely for A one-dimensional projector and for $A = \hat{N}$ the number operator. Levy's lemma reproduced a qualitatively correct scaling behavior with the dimension of the Hilbert, although with very small pre-factor as compared to the exact ones. The quadratic scaling with the error ϵ predicted by Levy's lemma was seen to reduce to linear scaling for sufficiently small ϵ in the cases studied. We have also noticed that quantum expectation of A can be regarded as a linear combination of random variables that decouple in the limit of large Hilbert space dimension. Here we limited ourselves to discuss a couple of examples of quantum operators whose expectation value fulfill a central limit type of result in such a limit i.e., a properly rescaled expectation becomes normally distributed.

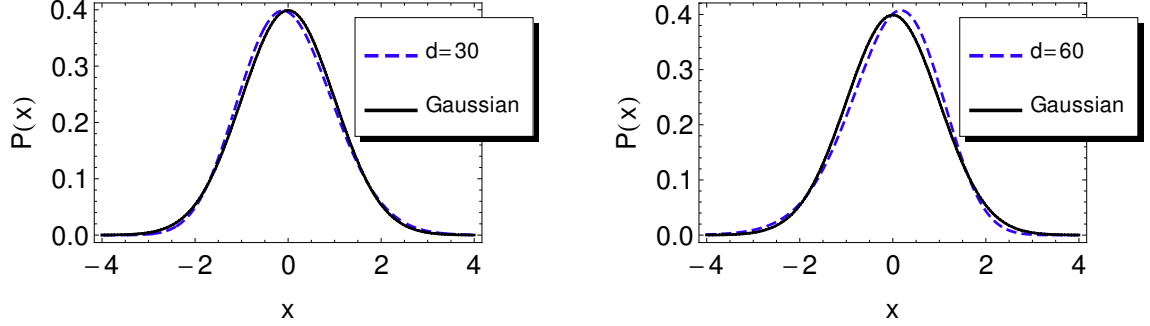


Figure 2. Approach to Gaussian and central limit theorem for the rescaled variable $\mathcal{Z} := (\mathcal{A} - \overline{\mathcal{A}}) / \sqrt{\kappa_2(\mathcal{A})}$. Left panel: the operator A has eigenvalues $a_k = k^2$. Right panel: $a_k = \ln(k)$. This latter case roughly corresponds to physical observables in the sense that the average value scales with the volume of the space, i.e. $\overline{\mathcal{A}} \sim \ln d \sim V$. The continuous line is the limiting case given by a standard normal distribution with zero mean and unit variance.

Before concluding it is important to mention that the Haar-induced measure over quantum pure states is well-known to be unphysical in different ways. In fact, sampling this measure with local quantum gates requires exponentially long random circuits [17]. Also, Haar typical quantum states are nearly maximally entangled [18], while low energy eigenstates of local quantum Hamiltonian fulfill area laws [19, 20, 21] i.e., they have low entanglement. In view of these remarks one may question the physical relevance of the results presented in this paper and look for more constrained prior measures over the $|\psi\rangle$'s. For example, in view of applications to foundations of statistical mechanics of closed systems, it would be interesting to generalize our results to “sections of constant energy” where one draws pure states uniformly under some constraint of the form $\langle \psi | H | \psi \rangle = E$ as e.g. in the spirit of [22].

A more ambitious goal would be to determine the distribution of expectations restricted to a set of “physical” states endowed with some “natural” measure. Instances of those ensembles are given in [24, 25, 26] where local random quantum circuits and matrix product states respectively have been considered. In these cases the lack of the maximal unitary invariance of the Haar measure represents the major obstruction one has to overcome.

Acknowledgement The authors would like to thank Karol Życzkowski for bringing to their attention references [5, 6, 23, 7].

This research is partially supported by the ARO MURI grant W911NF-11-1-0268 and NSF grants No. PHY-969969 and No. PHY-803304.

Appendix A. Some useful identities

Here we want to prove some identities which provide several important relations for the coefficients in Eq. (7) and eq. (12). We begin by considering the following function for y, Ω real $M(y, \Omega) := e^{\overline{(\mathcal{A} - \Omega)}y}$ which is somehow the moment generating function with a shift. The corresponding Gaussian integral is well defined when $y(\mathcal{A} - \Omega \mathbb{I}) < 0$. To satisfy this condition we assume $\Omega > \max\{a_j\}$ and $y > 0$.

The Gaussian integral converges and gives

$$M(y, \Omega) = \frac{(d-1)!}{(-i)^d} \int \frac{dr}{2\pi} \frac{e^{-ir}}{\prod_j (r - r_j)}, \quad r_j = iy(a_j - \Omega)$$

The integral can be evaluated again with complex integration. The contour integral must be closed in the lower half plane, all the poles are on the negative imaginary axis and we get

$$\begin{aligned} M(y, \Omega) &= \frac{(d-1)!}{(-i)^{d-1}} \sum_{k=1}^d \frac{e^{-ir_k}}{\prod_{j \neq k} (r_k - r_j)} \\ &= \frac{(d-1)!}{y^{d-1}} \sum_{k=1}^d \frac{e^{y(a_k - \Omega)}}{\prod_{j \neq k} (a_k - a_j)} \end{aligned} \quad (\text{A.1})$$

The same quantity at leading order in y can be computed by first expanding the integral around $y = 0$ and then integrating. We obtain then

$$M(y, \Omega) = 1 + O(y) \quad (\text{A.2})$$

Equating eqns. (A.1) and (A.2) term by term we arrive at

$$\sum_{k=1}^d \frac{(a_k - \Omega)^n}{\prod_{j \neq k} (a_k - a_j)} = \begin{cases} 0 & 0 \leq n \leq d-2 \\ 1 & n = d-1 \end{cases}. \quad (\text{A.3})$$

These equations have been obtained assuming $\Omega > \max\{a_j\}$, but since they are analytic in Ω they must be true for all complex Ω . In particular, setting $\Omega = 0$ in (A.3) we deduce that $\chi(\lambda)$ is regular at $\lambda = 0$, and in fact analytic and we obtain eq. (9). Considering instead $\Omega > \max\{a_k\}$ (resp. $\Omega < \min\{a_k\}$) and applying the result to eq. (12) we obtain that $P(x) = 0$ for $x > \max\{a_k\}$ (resp. $x < \min\{a_k\}$).

Appendix B. Degenerate case

We provide here some additional steps needed to obtain eq. (19). In practice we need to write down the differentiation in eq. (16). The first step is the following:

$$\partial_r^{(n_k-1)} [e^{-ir} g(r)] = \sum_{M=0}^{n_k-1} \binom{n_k-1}{M} (-i)^{n_k-1-M} e^{-ir} \partial_r^M g(r)$$

Then we need the multinomial formula

$$\partial_r^M \left[\prod_{j=1}^{\ell} g_j(r) \right] = \sum_{m_1=0}^M \cdots \sum_{m_{\ell}=0}^M \delta_{M, \sum_{j=1}^{\ell} m_j} M! \prod_{j=1}^{\ell} \frac{g_j^{(m_j)}(r)}{m_j!}$$

As customary in physics we write the multiple sum as

$$\sum_{m_1=0}^M \cdots \sum_{m_{\ell}=0}^M \delta_{M, \sum_{j=1}^{\ell} m_j} = \sum_{\sum_{j=1}^{\ell} m_j = M}^{\ell}.$$

Note that when applying this to eq. (16) we miss the term with $j = k$ so we have $\ell - 1$ (constrained) sums over m_i . The final bit is

$$\partial_r^{m_j} [(r - r_j)^{-n_j}] = (-1)^{m_j} \frac{(n_j + m_j - 1)!}{(n_j - 1)!} \frac{1}{(r - r_j)^{n_j + m_j}}.$$

Putting things together, the derivative reads

$$\begin{aligned} \partial_r^{(n_k-1)} \left[\frac{e^{-ir}}{\prod_{j \neq k}^\ell (r - r_j)^{n_j}} \right] = \\ \sum_{M=0}^{n_k-1} \binom{n_k-1}{M} (-i)^{n_k-1-M} e^{-ir} \times \\ \times \sum_{\sum_{j \neq k}^N m_j = M} M! \prod_{j \neq k} (-1)^{m_j} \frac{(n_j + m_j - 1)!}{m_j! (n_j - 1)!} \frac{1}{(r - r_j)^{n_j + m_j}} \end{aligned}$$

Going back we get

$$\begin{aligned} M(y, \Omega) = \frac{(d-1)!}{(-i)^{d-1}} \sum_k \frac{1}{(n_k-1)!} \times \\ \sum_{M_k=0}^{n_k-1} \binom{n_k-1}{M_k} (-i)^{n_k-1-M_k} e^{-ir_k} \times \\ \times \sum_{\sum_{j \neq k}^N m_j = M_k} M_k! \prod_{j \neq k} \frac{(n_j + m_j - 1)!}{m_j! (n_j - 1)!} \frac{(-1)^{m_j}}{(r_k - r_j)^{n_j + m_j}} \end{aligned}$$

Correctly all i factors cancel out and we are left with

$$M(y, \Omega) = (d-1)! \sum_{k=1}^\ell \sum_{M_k=0}^{n_k-1} \frac{(-1)^{M_k} e^{y(a_k - \Omega)}}{y^{d+M_k-n_k} (n_k-1-M_k)!} \beta_k(M_k) \quad (\text{B.1})$$

$$\beta_k(M_k) = \sum_{\sum_{j \neq k}^\ell m_j = M_k} \prod_{\substack{j=1 \\ j \neq k}}^\ell \frac{(n_j + m_j - 1)!}{m_j! (n_j - 1)!} \frac{1}{(a_k - a_j)^{n_j + m_j}} \quad (\text{B.2})$$

which correctly reduces to eq. (A.1) when all $n_k = 1$. On the other hand we still have $M(y, \Omega) = 1 + \frac{y}{d} \text{tr}(A - \Omega) + O(y^2)$ which can be used to show directly that $M(y, \Omega)$ is regular at $y = 0$ and hence analytic in the whole complex plane.

References

- [1] Terence Tao. *Topics in Random Matrix Theory*. American Mathematical Society, April 2012. 1
- [2] Noah Linden, Sandu Popescu, Anthony J. Short, and Andreas Winter. Quantum mechanical evolution towards thermal equilibrium. *Physical Review E*, 79(6):061103, June 2009. 1
- [3] Lorenzo Campos Venuti, N. Tobias Jacobson, Siddhartha Santra, and Paolo Zanardi. Exact Infinite-Time statistics of the loschmidt echo for a quantum quench. *Physical Review Letters*, 107(1):010403, July 2011. 1
- [4] Sandu Popescu, Anthony J. Short, and Andreas Winter. Entanglement and the foundations of statistical mechanics. *Nat Phys*, 2(11):754–758, November 2006. 1, 4.1
- [5] Charles F. Dunkl, Piotr Gawron, John A. Holbrook, Zbigniew Puchała, and Karol Życzkowski. Numerical shadows: Measures and densities on the numerical range. *Linear Algebra and its Applications*, 434(9):2042–2080, May 2011. 1, 6
- [6] Charles F Dunkl, Piotr Gawron, John A Holbrook, Jarosław A Miszczak, Zbigniew Puchała, and Karol Życzkowski. Numerical shadow and geometry of quantum states. *Journal of Physics A: Mathematical and Theoretical*, 44(33):335301, August 2011. 1, 6
- [7] Thierry Gallay and Denis Serre. Numerical measure of a complex matrix. *Communications on Pure and Applied Mathematics*, 65(3):287–336, March 2012. 1, 5, 6

- [8] Hermann Weyl. *The classical groups: their invariants and representations*. Princeton University Press, 1997. 2
- [9] Roe Goodman and Nolan R. Wallach. *Representations and invariants of the classical groups*. Cambridge University Press, January 2000. 2
- [10] Pierre-Yves Gaillard. Around the chinese remainder theorem. <http://www.iecn.u-nancy.fr/~gaillap/DIVERS/Chinese.Remainder.Theorem/>, 2008. 3
- [11] John Preskill. Course information for physics, computer science, quantum computation. <http://theory.caltech.edu/~preskill/ph229/>, 2004. 3.1
- [12] J. v. Neumann. Beweis des ergodensatzes und des H-Theorems in der neuen mechanik. *Zeitschrift für Physik*, 57(1-2):30–70, January 1929. 3.1
- [13] J. Neumann. Proof of the ergodic theorem and the h-theorem in quantum mechanics. *The European Physical Journal H*, 35(2):201–237, September 2010. 3.1
- [14] Michel Ledoux. *The Concentration of Measure Phenomenon*. American Mathematical Soc., August 2001. 4
- [15] Sandu Popescu, Anthony J Short, and Andreas Winter. The foundations of statistical mechanics from entanglement: Individual states vs. averages. *arXiv:quant-ph/0511225*, November 2005. 4.1
- [16] Assa Auerbach. *Interacting electrons and quantum magnetism*. Springer, 1994. 5
- [17] Joseph Emerson, Yaakov S Weinstein, Marcos Saraceno, Seth Lloyd, and David G Cory. Pseudo-Random unitary operators for quantum information processing. *Science*, 302(5653):2098–2100, December 2003. 6
- [18] Patrick Hayden, Debbie W. Leung, and Andreas Winter. Aspects of generic entanglement. *Communications in Mathematical Physics*, 265(1):95–117, March 2006. 6
- [19] J. Eisert, M. Cramer, and M. B. Plenio. Colloquium: Area laws for the entanglement entropy. *Reviews of Modern Physics*, 82(1):277–306, February 2010. 6
- [20] Alioscia Hamma, Radu Ionicioiu, and Paolo Zanardi. Ground state entanglement and geometric entropy in the kitaev model. *Physics Letters A*, 337(1-2):22–28, March 2005. 6
- [21] Alioscia Hamma, Radu Ionicioiu, and Paolo Zanardi. Bipartite entanglement and entropic boundary law in lattice spin systems. *Physical Review A*, 71(2):022315, February 2005. 6
- [22] Markus P. Müller, David Gross, and Jens Eisert. Concentration of measure for quantum states with a fixed expectation value. *Communications in Mathematical Physics*, 303(3):785–824, March 2011. 6
- [23] Zbigniew Puchała, Jarosław Adam Miszczak, Piotr Gawron, Charles F Dunkl, John A Holbrook, and Karol Życzkowski. Restricted numerical shadow and geometry of quantum entanglement. *arXiv:1201.2524*, January 2012. 1, 6
- [24] Alioscia Hamma, Siddhartha Santra, and Paolo Zanardi. Quantum entanglement in random physical states. *arXiv:1109.4391*, September 2011. 6
- [25] Silvano Garnerone, Thiago R. de Oliveira, and Paolo Zanardi. Typicality in random matrix product states. *Physical Review A*, 81(3):032336, March 2010. 6
- [26] Silvano Garnerone, Thiago R. de Oliveira, Stephan Haas, and Paolo Zanardi. Statistical properties of random matrix product states. *Physical Review A*, 82(5):052312, November 2010. 6